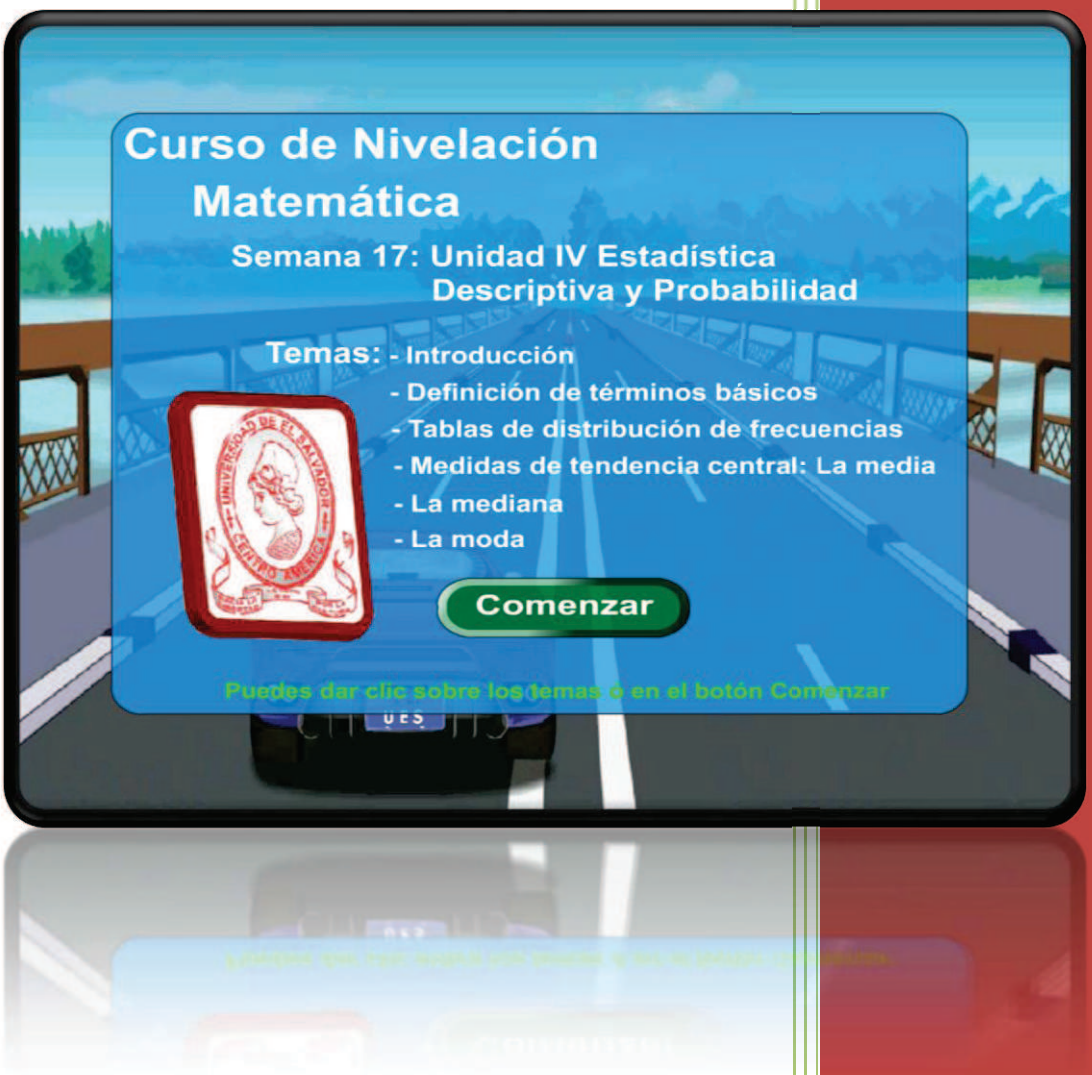
UNIVERSIDAD DE EL SALVADOR
Curso de Nivelación 2014

MATEMÁTICA



Universidad de El Salvador, Derechos Reservados

CURSO DE MATEMÁTICA EN LÍNEA

Contenido

INTRODUCCION..... 2

DEFINICIÓN DE TÉRMINOS BÁSICOS..... 3

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS..... 5

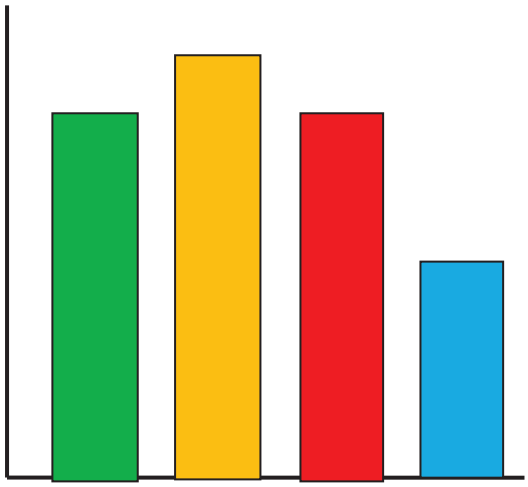
MEDIDAS DE TENDENCIA CENTRAL..... 8

LA MEDIA..... 8

LA MEDIANA..... 10

LA MODA..... 12

Universidad de El Salvador, Derechos Reservados 2013



4

Estadística descriptiva y probabilidad

Secciones

Introducción
4.1 Medidas numéricas
4.2 Probabilidad
4.3 Técnicas de conteo y probabilidad

Universidad de El Salvador, Derechos Reservados 2013

INTRODUCCIÓN

La estadística estudia los métodos para recolectar, resumir y analizar datos, así como también sacar conclusiones y tomar decisiones apropiadas basadas en el análisis.

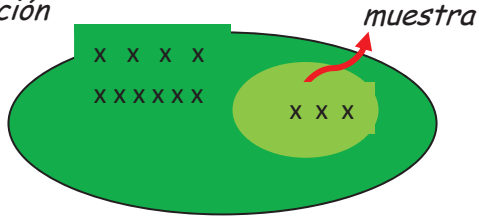
4.1 Medidas numéricas

Si bien no hay definición exacta de estadística, se puede decir que la *"estadística es el estudio de los métodos y procedimientos para recolectar, resumir y analizar datos y para hacer inferencias científicas partiendo de tales datos".*

La colección, o conjunto, de individuos, objetos, elementos o eventos cuyas propiedades serán analizadas se llama ***población.***

Podemos distinguir entre poblaciones tangibles y poblaciones conceptuales. Una **población es tangible** si consta de elementos físicos reales que forman un conjunto finito. Por ejemplo, si estamos considerando el estudio de la altura de los alumnos de una clase, el conjunto de estos alumnos es una población tangible. Una **población conceptual** no tiene elementos reales, sino que sus casos se obtienen por la repetición de un experimento. Por ejemplo, las pruebas de resistencia a las que un ingeniero somete a distintas muestras de asfalto.

población *muestra*



Muestra *es un subconjunto de la población.*

Cuando se desea conocer alguna característica de una población y muestra se lleva acabo un ***censo*** y una ***encuesta*** respectivamente.

Universidad de El Salvador, Derechos Reservados 2013

Definición de términos básicos

La característica de interés sobre cada elemento individual de una población o muestra se llama ***variable.*** La edad de un estudiante que ingresa a la universidad, el departamento de procedencia, su estatura y su peso son cuatro variables.

El *valor de la variable que corresponde a un elemento de una población o muestra* se llama ***dato.*** Este valor puede ser un número, una palabra o un símbolo. Por ejemplo, Manuel Petronilo Osegueda ingresó a la universidad a la edad de "18" años, procede del departamento de La Libertad, mide "1.70 mt" y pesa "80 kg". Estas cuatro partes de datos son los valores de las cuatro variables aplicadas a Manuel Petronilo Osegueda.

El *conjunto de valores recolectados para la variable de cada uno de los elementos que pertenecen a la muestra* se llaman ***datos.*** Por ejemplo, el conjunto de 25 estaturas recolectadas de 25 estudiantes.

Atendiendo a su naturaleza, las variables bajo estudio pueden clasificarse según los siguientes tipos:

- *Variables* ***cualitativas*** *son aquellas que no toman valores numéricos*. Por ejemplo, el color de una tela o la marca de un televisor.
- *Variables* ***cuantitativas*** *son aquellas que toman valores numéricos*. A su vez, éstas pueden clasificarse en **discretas** y **continuas.**
  - *Variables* ***discretas*** *son las que asumen solo valores, por lo general enteros.* Por ejemplo, el número N de alumnos en una asignatura.
  - *Variables* ***continuas*** *son aquellas que pueden asumir virtualmente cualquier valor en un determinado intervalo real*. Por ejemplo, el peso *x* o la altura *h* de un estudiante.
- *Variables* ***jerarquizadas*** *son las que comúnmente se refieren a las evaluaciones subjetivas cuando los conceptos se jerarquizan según la preferencia o logro.* Por ejemplo, en los concursos, los lugares se jerarquizan en 1er. lugar, 2º. lugar, etc.; las posiciones de los equipos se les asignan los números 1, 2, 3,…

Un ***experimento*** *es una actividad planeada cuyos resultados proporcionan un conjunto de datos.* El experimento incluye las actividades para seleccionar los elementos y obtener los valores de los datos.

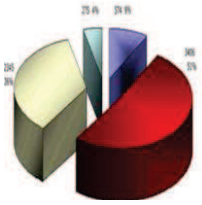
Universidad de El Salvador, Derechos Reservados 2013

El *valor numérico que resume todos los datos de una población completa* se llama ***parámetro***. La "proporción" de estudiantes que tenían más de 20 años de edad cuando ingresaron a la universidad es un ejemplo de parámetro. Un parámetro es un valor que describe a la población. Nos auxiliamos de letras griegas a menudo para denotar a un parámetro, por ejemplo los símbolos: μ, σ^2, son la *media* y *varianza* poblacional respectivamente.

El ***estadístico*** *es el valor numérico que resume los datos de la muestra.* La estatura "media" encontrada al utilizar el conjunto de 25 estaturas es un ejemplo de un estadístico. Casi todos los estadísticos se determinan con ayuda de fórmulas y suelen simbolizarse auxiliándose de letras del alfabeto español: $\bar{x}$ y s^2, son la *media* y *varianza* muestral respectivamente.

Clasificación de la estadística

La estadística suele dividirse en **estadística descriptiva, probabilidad** y **estadística inferencial**.



*La **estadística descriptiva** consiste en los procedimientos utilizados para resumir y describir las características importantes de un conjunto de datos.*

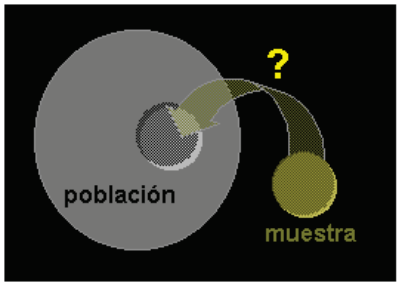
La estadística descriptiva utiliza ***métodos gráficos*** y ***numéricos*** para describir un conjunto de datos.

*La **probabilidad** estudia las leyes que rigen los fenómenos aleatorios. Es el modelo matemático formal de la estadística.*

Si el conjunto de datos es la población entera, sólo se necesita sacar conclusiones con base en la estadística descriptiva. Sin embargo, podría ser demasiado caro o tardado enumerar toda la población. Quizá enumerar la población la destruiría, como en el caso de la prueba "resistencia a la ruptura" de una pieza de un lote. Por éstas u otras razones, es preferirle seleccionar en algunos casos sólo una muestra de la población. Al examinar la muestra, se

Universidad de El Salvador, Derechos Reservados 2013

desean contestar las preguntas acerca de la población como un todo. La rama de la estadística que trata con este problema se llama **estadística inferencial.**



*La **estadística inferencial** consiste en los procedimientos utilizados para hacer inferencias acerca de las características de la población, a partir de la información contenida en una muestra extraída de esta población.*

Tablas de distribución de frecuencias

Cuando un gran número de datos se presenta, la presentación con la palabra escrita se vuelve ineficiente y pesada. En este caso conviene usar cuadros y gráficas. Si los datos estadísticos se presentan en forma de cuadros o tablas, los datos son arreglados sistemáticamente en columnas.

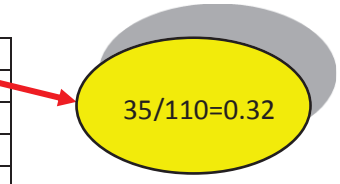
Una tabla de distribución de frecuencias (f), es un cuadro que contiene información resumida.Si en una tabla de distribución de frecuencias, en lugar de dar el número de veces nos fijamos en la proporción de veces, tenemos la distribución de frecuencias relativas (fr).

Ejemplo 4.1.

Para conocer la ansiedad por trabajar en la computadora se aplicó a 100 estudiantes de un curso en el que se usó computadora un cuestionario. Una de las preguntas de dicho cuestionario era: “¿me gusta usar las computadoras?”. Las respuestas a esta pregunta en particular fueron:

Universidad de El Salvador, Derechos Reservados 2013

Respuesta	f	f_r
Totalmente de acuerdo	35	0.32
De acuerdo	58	0.53
Ligeramente de acuerdo	8	0.07
Ligeramente en desacuerdo	2	0.02
En desacuerdo	1	0.01
Totalmente en desacuerdo	6	0.05
	$n=\Sigma f=110$	$\Sigma f_r=1$



Donde Σf significa sumar todas las frecuencias y de forma parecida Σf_r.

Cuando el número de valores que toma la variable es elevado, se usa el **método de las frecuencias agrupadas.** Aquí los datos se distribuyen en **clases** o **categorías** y luego se determina el número de datos pertenecientes a cada clase, que es la **frecuencia de clase, *f.*** Un ordenamiento de esta naturaleza se conoce como una **distribución de frecuencias** o **tabla de frecuencias.**

Ejemplo 4.2.

Dado los siguientes puntajes de un examen de estadística, construya una tabla de distribución de frecuencias.

5.8	4.5	8.0	9.3	8.6	7.0	6.5	6.4	6.6	9.6
8.8	7.5	8.4	5.6	6.2	9.3	7.2	7.0	8.6	7.2
7.5	3.7	8.8	6.1	6.6	9.5	6.8	6.2	6.8	6.8
5.6	7.6	8.7	4.2	5.3	8.3	8.0	8.1	7.0	7.5
7.0	8.4	4.8	9.2	9.0	7.8	8.9	7.3	7.4	7.6

Procedimiento

1. Identifique los puntajes máximo y mínimo (*Máx*=9.6, *Mín*=3.7) y determine el Rango, ***R,*** donde *R*= *Rango*=*Máx* – *Mín*=9.6 – 3.7=5.9.

2. Decidir el número de clases, ***h***, que se va usar. No existe número idóneo, aunque conviene evitar los extremos. Puede aconsejarse un número discreto entre 5 y 20 clases. Utilizaremos aquí, para el número de clases, el método de la raíz cuadrada del número de datos. En el caso nuestro $\sqrt{50}$ = 7.07, que se redondea a 8. Así, pues *h*=8.

Universidad de El Salvador, Derechos Reservados 2013

3. Dividir el rango *R* entre *h*, para obtener la **amplitud** o **ancho de clase, *c*.**
   Aquí, *c=R/h*= 5.9/8=0.7375, el cual redondeamos a *c*=0.74, ya que por regla *mantenemos un sitio decimal más del que hay en la información original*. Si el *Máx* hubiese caído fuera de la última clase, tomaríamos 0.75 si es necesario.

4. El primer intervalo se forma agregando la amplitud 0.74 al *Mín*=3.7, siendo dicho intervalo o clase 3.7≤x<4.44, [3.7, 4.44[ó bien con la notación 3.7 - < 4.44. La segunda clase es 4.44 - < 5.18, etc., hasta obtener finalmente la siguiente tabla:

Intervalo
3.7 - < 4.44
4.44 - < 5.18
5.18 - < 5.92
5.92 - < 6.66
6.66 - < 7.40
7.40 - < 8.14
8.14 - < 8.88
8.88 - < 9.62

5. Encontrar las frecuencias de clase, *f*, contando el número de datos que caen en cada intervalo y este número anotarlo en la columna correspondiente a las frecuencias. El total de esta columna debe ser igual a *50 (n),* el número de datos.

Puntajes	f
3.70 - < 4.44	2
4.44 - < 5.18	2
5.18 - < 5.92	4
5.92 - < 6.66	6
6.66 - < 7.40	10
7.40 - < 8.14	10
8.14 - < 8.88	9
8.88 - < 9.62	7
Total	n=Σf=50

*Se denomina **frecuencia acumulada ascendente, F_a**, hasta un intervalo de clase, a la suma de las frecuencias de los intervalos anteriores a él, hasta los de ese intervalo.*

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.3.
Considérese la tabla de distribución de frecuencias del ejemplo 4.2 y agregue la columna de frecuencia acumulada ascendente.

Intervalo	f	F_a
3.70 - < 4.44	2	2
4.44 - < 5.18	2	4
5.18 - < 5.92	4	8
5.92 - < 6.66	6	14
6.66 - < 7.40	10	24
7.40 - < 8.14	10	34
8.14 - < 8.88	9	43
8.88 - < 9.62	7	50

Medidas de tendencia central

Un **promedio** es un valor que es típico, o representativo de un conjunto de datos. Como tales valores tienden a situarse en el centro del conjunto de datos ordenados según su magnitud, los promedios se conocen también como **medidas de centralización.** Se pueden definir varias medidas de centralización, las más comunes son la **media aritmética** o brevemente **media,** la **mediana** y la **moda.** Cada una de ellas tiene sus ventajas y desventajas, dependiendo la aplicación de una o de otra, de los resultados que se pretendan sacar de los datos.

La media: $\bar{x}$

*La **media aritmética** o **media** de un conjunto de n números $x_1, x_2, \ldots x_n$ se representa por $\bar{x}$ y se define como*

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Para representar la suma $x_1 + x_2 + \ldots + x_n$ se usa el símbolo $\sum$. Así, $\sum_{i=1}^{n} x_i$ significa que debemos sumar los elementos x_i, los que inician con x_1 (cuando i=1), x_2 (cuando i=2) y así sucesivamente hasta finalizar con x_n (cuando i=n). Con esta notación, la media se escribe de la manera siguiente:

Universidad de El Salvador, Derechos Reservados 2013

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

La media aritmética de una población se denota por: μ.

Cuando se redondea una respuesta, se acordará mantener un sitio decimal más en la respuesta del que hay en la información original. Para evitar la acumulación de redondeo, éste sólo se hará en la respuesta final y no en los pasos intermedios.

Ejemplo 4.4.
Cinco piezas de una aleación de níquel, seleccionadas al azar, tienen las siguientes cantidades de níquel (en porcentaje): 78.7, 79.4, 79.5, 78.9, 79.7 y 78.8. La media aritmética es

$$\bar{x} = \frac{78.7 + 79.4 + 79.5 + 78.9 + 79.7 + 78.8}{5} = \frac{475}{6} = 79.17\ \%$$

La respuesta se expresa con un decimal más del que aparece en los datos, redondeamos entonces hasta las centésimas

Si los números $x_1, x_2, \ldots, x_k$ se presentan $f_1, f_2, \ldots, f_k$ veces, respectivamente (es decir, se presentan con frecuencias $f_1, f_2, \ldots, f_k$), la media aritmética es

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots + f_kx_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum f_ix_i}{\sum f_i} = \frac{\sum f_ix_i}{n} \quad (A)$$

donde $n = \sum f_i$ es la **frecuencia total** es decir, el número total de datos.

Ejemplo 4.5.
Si 4, 8, 9 y 6 se presentan con frecuencias 3, 5, 4 y 2 respectivamente, la media aritmética es

$$\bar{x} = \frac{(3)(4) + (5)(8) + (4)(9) + (2)(6)}{3 + 5 + 4 + 2} = 7.1$$

A veces se asocia a los números $x_1, x_2, \ldots, x_k$ ciertos **factores** o **pesos** $w_1, w_2, \ldots, w_k$ que dependen de la significación o importancia de cada uno de los números. En este caso

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \ldots + w_kx_k}{w_1 + w_2 + \ldots + w_k} = \frac{\sum wk}{\sum w}$$

se llama **media aritmética ponderada.**

Universidad de El Salvador, Derechos Reservados 2013

Nótese la similitud con la fórmula (A), que puede considerarse como una media aritmética con los pesos $f_1, f_2, \ldots, f_k$.

Ejemplo 4.6.

Un estudiante de la asignatura de Estadística obtuvo en sus cinco evaluaciones: 6.0, 8.0, 5.4, 7.0 y 6.3. Hallar la nota final de ciclo si las ponderaciones son: 0.25, 0.10, 0.30, 0.10 y 0.25 respectivamente.

$$\bar{x} = \frac{(6.0)(0.25)+(8.0)(0.10)+(5.4)(0.30)+(7.0)(0.10)+(6.3)(0.25)}{0.25+0.10+0.30+0.10+0.25} = 6.20$$

Media aritmética calculada a partir de datos agrupados

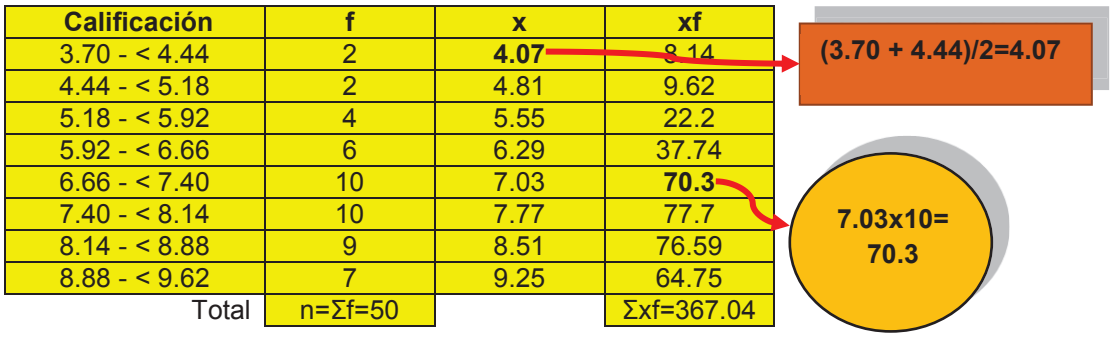
Para aproximar la media de datos organizados en una distribución de frecuencias, se empezará por asumir que las observaciones de cada clase están representadas por la **marca de clase** o **punto medio** x_i (ó P_m) y f_i su correspondiente frecuencia de clase. Con esto la fórmula (A) se modifica a

$$\bar{x} = \frac{\sum x_i f_i}{n}$$

Ejemplo 4.7.

Considérese la tabla de distribución de frecuencias del ejemplo 4.2 y calcule la media.

Calificación	f	x	xf
3.70 - < 4.44	2	**4.07**	8.14
4.44 - < 5.18	2	4.81	9.62
5.18 - < 5.92	4	5.55	22.2
5.92 - < 6.66	6	6.29	37.74
6.66 - < 7.40	10	7.03	**70.3**
7.40 - < 8.14	10	7.77	77.7
8.14 - < 8.88	9	8.51	76.59
8.88 - < 9.62	7	9.25	64.75
Total	n=Σf=50		Σxf=367.04



La media es $\bar{x}$ =Σxf / n =367.04/50=7.341.

La desventaja de la media es que es afectada por valores extremos

La mediana: $\tilde{x}$

*La **mediana** de una colección de datos ordenados de menor a mayor es el valor medio o la media aritmética de los dos valores medios.*

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.8.
Hallar la mediana de los números: 8, 3, 8, 4, 6, 5, 8, 10, 4.
Ordenando los datos de acuerdo a su magnitud: 3, 4, 4, 5, 6, 8, 8, 8, 10, se tiene que $\tilde{x}$ =6.

Ejemplo 4.9.
Hallar la mediana de los números: 9, 2, 7, 11, 14, 6.
Ordenando de acuerdo a su magnitud: 2, 6, 7, 9, 11, 14, la mediana es el valor medio entre 7 y 9, el cual es $\tilde{x}$ =(7 + 9)/2 = 8.0.

Si se ordenan los datos de menor a mayor y si *n* es impar, la mediana es el dato que aparece en el lugar *(n+1)/2*; si *n* es par, la mediana es el promedio de los datos que están en los lugares *n/2* y *(n/2) + 1*. Así**:**

si *n* es impar $\tilde{x} = X_{(n+1)/2}$ y si *n* es par $\tilde{x} = [X_{n/2} + X_{(n/2)+1}] / 2.$

La mediana calculada a partir de datos agrupados

Supóngase que las observaciones (datos) en cada clase se "dispersan uniformemente" a través del intervalo de clase, la mediana se obtiene por interpolación por medio de la fórmula

$$\tilde{x} = l_i + c\left[\frac{n/2 - F_a}{f}\right]$$

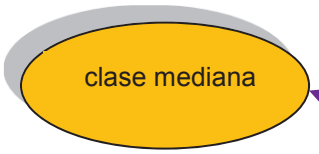
donde l_i: límite inferior de la clase mediana
c: ancho de la clase mediana
n: número de datos
F_a: frecuencia acumulada anterior a la clase mediana
f: frecuencia de la clase mediana

Ejemplo 4.10.
La edad de cada uno de los 50 visitantes que asistieron a una exhibición científica viene dada por la siguiente tabla de frecuencias. Calcule la mediana.

Edad	f
0 - < 10	6
10 - <20	18
20 - < 30	11
30 - < 40	3
40 - < 50	0
50 - < 60	8
60 - < 70	4

Universidad de El Salvador, Derechos Reservados 2013

Completando la columna correspondiente a la frecuencia acumulada ascendente se tiene:



Edad	f	Fa
0 - < 10	6	6
10 - <20	18	24
20 - < 30	**11**	35
30 - < 40	3	38
40 - < 50	0	38
50 - < 60	8	46
60 - < 70	4	50

Aquí, $n/2$ =50/2=25, entonces ¿cuál es la primer frecuencia acumulada ascendente que contiene a 25?, *¡35!*, entonces la clase mediana es 20 - < 30, de donde l_i=**20**, F_a=**24** y f=**11**.

Así, al utilizar la fórmula se obtiene

$$\tilde{x} = l_i + c\left[\frac{n/2 - F_a}{f}\right]$$

$$\tilde{x} = \mathbf{20} + 10\left[\frac{25 - \mathbf{24}}{\mathbf{11}}\right] = 20.9 \text{ años.}$$

La moda: $\hat{x}$

*La **moda** de una serie de datos es aquel dato que se presenta con la mayor frecuencia, es decir, es el valor más común.*

La moda puede no existir, incluso si existe puede no ser única.

Ejemplo 4.11.

Considérense los datos 9, 2, 7, 11, 14, 7, 2, 7. El valor 7 ocurre tres veces, el 2 aparece dos veces y el resto ocurre una vez cada uno, luego, $\hat{x} = 7$.

Universidad de El Salvador, Derechos Reservados 2013

La moda no es una medida de tendencia central muy usual, pero sirve en los negocios, por ejemplo para determinar qué tamaño del producto es el de mayor demanda. Por ejemplo, un fabricante de camisetas está interesado en la talla de mayor demanda. Una distribución que tiene una sola moda se llama **unimodal.**

Moda calculada a partir de datos agrupados

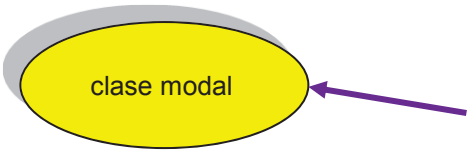
En el caso de datos agrupados donde se ha construido una curva de frecuencias para ajustar los datos, la moda será el valor (o valores) de x correspondientes al máximo (o máximos) de la curva.

Dada una distribución de frecuencias o un histograma, la moda puede calcularse por la fórmula

$$\hat{x} = l_i + c\left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right]$$

donde
l_i: límite inferior de la clase modal
c: ancho de la clase modal
$\Delta_1 = f_i - f_{i-1}$: exceso de la frecuencia modal sobre la frecuencia de la clase contigua inferior
$\Delta_2 = f_i - f_{i+1}$: exceso de la frecuencia modal sobre la frecuencia de la clase contigua superior

Ejemplo 4.12. Determine la moda del ejemplo 4.10.



Edad	f
0 - < 10	6
10 - <20	18
20 - < 30	11
30 - < 40	3
40 - < 50	0
50 - < 60	8
60 - < 70	4

Aquí, $l_i = 10$, $\Delta_1 = f_i - f_{i-1} = 18 - 6 = 12$, $\Delta_2 = f_i - f_{i+1} = 18 - 11 = 7$

y $c = 10$. Entonces

$$\hat{x} = l_i + c\left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right]$$

$$\hat{x} = 10 + 10\left[\frac{12}{12 + 7}\right]$$

$$\hat{x} = 16.3 \text{ años}$$

Universidad de El Salvador, Derechos Reservados 2013